コメントデータを用いた動画内特定印象シーン分析手法の改良

笹原　彰斗† Cheng Asaka Lan†† 山口　史弥†† 上田真由美†††,†††† 中島　伸介††

† 京都産業大学 先端情報学研究科　〒 603–8555 京都府京都市北区上賀茂本山
†† 京都産業大学 情報理工学部　〒 603–8555 京都府京都市北区上賀茂本山
††† 追手門学院大学 経営学部　〒 567–8620 大阪府茨木市太田東芝町 1 − 1
†††† 大阪大学 D3 センター　〒 567–0047 大阪府茨木市美穂ケ丘 5-1
E-mail: {i2486093, a007532, yamaguchi.fumiya, nakajima}@cc.kyoto-su.ac.jp, may-ueda@otemon.ac.jp

あらまし　近年，動画共有サイトには膨大な数の動画が投稿されている．多くのユーザはタイトルやサムネイルから好みの動画を判断して視聴するが，既存の情報では実際に再生するまでユーザが真に求めている動画か判断することが困難である．そこで我々は，「ユーザが動画を視聴して直感的に感じる印象」を評価項目別スコアとして算出する直感的動画検索システムの開発に取り組んできた．また，ユーザが求める評価項目に該当するシーンが動画内のどの時点で現れるかを特定するため，ニコニコ動画のタイムスタンプ付きコメントデータを用いた時系列印象分析手法を提案してきた．しかし，従来手法は分類モデル構築手法や分析アルゴリズムにおけるパラメータ設定の面で改善の余地があり，シーン検出精度に課題が残されていた．そこで本稿では，従来手法に対してモデル構築手法の改良，分析アルゴリズムにおけるパラメータの組み合わせ探索を行い，シーン検出精度の向上を図る．

キーワード　動画分析，コメント分析，機械学習

1 はじめに

近年，動画共有サイトには膨大な数の動画が投稿されており，多くのユーザはキーワード検索を行った後，タイトルやサムネイルから好みの動画を選択し視聴している．しかしながら，ユーザが直感的に視聴したいと感じる動画を探索することは，従来のキーワード検索のみでは困難である．例えば，ユーザが「エモい動画」を視聴したい場合，従来のシステムにおいて「エモい」というキーワードで検索を行えば，タイトルや説明文に当該語句を含む動画が検索結果として提示される．しかし，メタデータにキーワードが含まれていても，視聴者が実際にその印象を受けるとは限らない．一方で，キーワードが含まれていなくても，当該印象を与える動画は多数存在する． そこで我々は，ユーザの直感的な動画探索を可能とする，評価項目別スコアを用いた直感的動画検索システム [1] を提案した．本システムは，動画の印象を表す 8 つの評価項目ごとに分類モデルを構築し，その特徴をレーダーチャートで提示するものである．これにより，タイトルや説明文に検索語句を含まない動画に対しても有効な検索が可能となる． 本稿では，以後「エモい」や「きゅん」といった項目を「印象項目」，印象項目の出現箇所を「印象シーン」と定義する．

しかし，直感的動画検索システムの実用化においては，動画全体のスコアだけでなく，印象シーンが時系列上のどの時点で現れるのかを特定することが重要である．動画全体のスコアが適切であっても，具体的な出現箇所が不明であれば，ユーザは目的の印象シーン（例：エモいと感じるシーン）を探し出して視聴するために動画全体を視聴する必要があり，それには多くの時間を要するためである．

この課題を解決するために，我々はこれまでに時系列印象分析手法を検討してきた．直感的動画検索システムの開発当初は YouTube を対象としていたが，特定のシーンに対する視聴者の反応（タイムスタンプ情報）を網羅的に取得することが仕様上困難であった．そこで本研究では，コメントが動画の再生時間と紐付けられて投稿・表示されるニコニコ動画へ分析対象を移行し，タイムスタンプ付きコメントデータを用いた時系列印象分析手法の開発を行っている．しかし，従来手法は分析アルゴリズムやパラメータ設定の面で改善の余地があり，シーン検出の精度に課題が残されていた [2]．そこで本稿では，従来手法に対してモデル構築手法の改良，分析アルゴリズムにおけるパラメータの組み合わせ探索を行い，シーン検出精度の向上を図る．

本稿の構成は以下の通りである．2 章では，関連研究について概説し本研究の位置付けを明確にする．3 章では，本研究の基盤となる直感的動画検索システムの概要および課題について述べる．4 章では，これまでに構築したモデルの分類精度向上のための改良手法について述べる．5 章では，構築したモデルを用い，動画内の印象シーンを特定する時系列分析手法について提案する．6 章では，提案手法の有効性を検証するための評価実験とその結果について述べる．最後に，第 7 章で本稿のまとめと今後の展望について述べる．

2 関連研究

本章では，動画推薦システム，感性分析，および時系列分析に関する関連研究について概説し，それらと本研究との差異について述べる．

2.1 動画推薦・検索システム

動画共有サイトにおけるコンテンツ発見を支援するため，多くの推薦システムが提案されている．従来の動画推薦では，協調フィルタリングやコンテンツベースフィルタリングが主流である．例えば，Hofmann ら [3] や加藤ら [4] は，ユーザのクリック履歴や検索ログから選好を推測する手法を検討している．近年では，Gheewala ら [5] や Sharma ら [6] のように，深層学習やグラフニューラルネットワークを用いてユーザとコンテンツの複雑な相互作用をモデル化する研究も進められている．しかし，これらの手法の多くはメタデータ（タイトル・タグ）や過去の行動履歴に依存しており，ユーザがその瞬間に抱く「直感的な感情」や「動画内の具体的な見どころ」を直接的に反映した検索・推薦には至っていない．

2.2 コメントを用いた感情・感性分析

ユーザコメントは，動画に対する視聴者の反応を直接的に表す情報源として広く分析されている．村上ら [7] は，日本のネットスラングである「w」に着目し，コメント量から動画の「面白さ」を定量化する手法を提案した．中村ら [8] は，感情表現辞書を構築し，「喜び」「悲しみ」「驚き」といった感情カテゴリに基づく動画検索手法を提案している．また，Siersdorfer ら [9] はサポートベクターマシン（SVM）を，Boddapati ら [10] は辞書ベースの手法を用いることで，YouTube 上のコメントを肯定・否定・中立に分類し，動画の評価予測を行った．さらに，Muszynski ら [11] は，テキストだけでなく音声や映像特徴を組み合わせたマルチモーダルな感情分析手法を提案している．本研究では，単なるポジティブ・ネガティブ分類ではなく，「エモい」「きゅん」といった動画に特化した印象項目に基づいた詳細な分析を行う．

2.3 時系列分析と本研究の位置付け

動画の時間軸に沿った視聴者の反応を捉える時系列分析は，動画検索において重要なテーマである．例えば，尾崎ら [12] は，コメントの投稿密度から行列を生成し，視聴者の「盛り上がり」を可視化するシステムを提案している．また，Rath ら [13] は BERT を用いて YouTube ライブのコメントに対する感情分析を行っている．さらに，Barakat ら [14] は，音声認識を用いて動画の音声をテキスト化し，時間経過に伴う感情の変化を捉える手法を提案している．

これらの先行研究は，コメントの投稿数に基づく「量的な盛り上がり」や，ポジティブ・ネガティブといった一般的な感性や極性の分析に留まっており，「エモい」や「きゅん」といった，ユーザが直感的に抱く具体的なニュアンスを持つ印象までは識別されていない．

そこで本研究では，時系列分析におけるシーン特定精度のさらなる向上を目的とする．具体的には，ニコニコ動画において頻出するノイズに対する前処理の導入，本タスクに適切な事前学習済みモデルの選定，そして時系列データから印象シーンを判定する分析アルゴリズムにおけるパラメータ探索を実施することで，従来手法よりも高精度な印象シーンの出現箇所分析を目指す．

3 直感的動画検索システム

本章では，本研究の基盤となる「直感的動画検索システム」[1] について述べる．

3.1 直感的動画検索システムの概要

開発当初の直感的動画検索システムは，YouTube に投稿された動画に対し，ユーザが付与したコメントデータを分析することで，動画が持つ「印象」を定量化し，提示するシステムである．本システムで採用している評価項目は，「エモい」「チル」「きゅん」「ぴえん」「映え」「イラつく」「萌え」「じわる」の 8 項目である．これらの語句は，従来の単純な極性分類では捉えきれない，動画特有の感性や雰囲気を反映している．

直感的動画検索システムでは，各評価項目に対応するモデルを構築し，コメント内容に基づいて動画ごとの評価項目別スコアを算出する．図 1 にシステムの利用イメージを示す．算出されたスコアをレーダーチャート等の形式でユーザに提示することで，ユーザは再生前に動画の雰囲気や評価項目の度合いを直感的に把握することが可能となる．

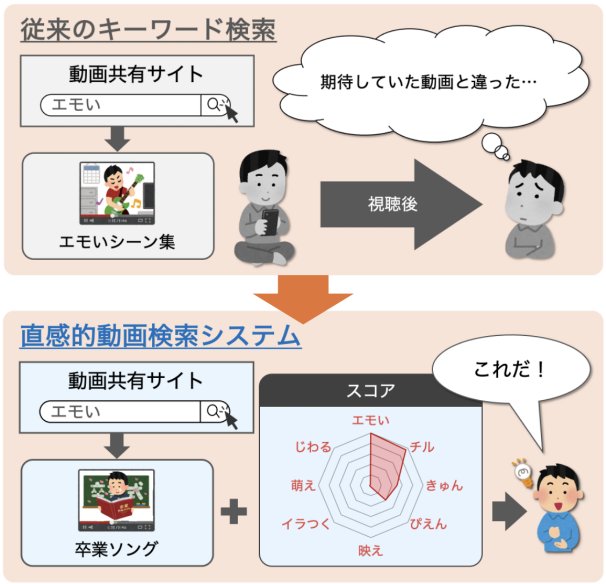

図 1 直感的動画検索システム（DEIM2022）

3.2 直感的動画検索システムの課題

前節で述べた直感的動画検索システムは，動画全体に対するスコア算出は可能であるものの，印象シーンの具体的な出現位置までは特定できない．そのため，動画全体のスコアが高くても，ユーザが求めている「エモいシーン」や「きゅんとするシーン」がどこにあるかが不明であれば，ユーザは目的のシーンを探し出して視聴するために動画全体を視聴せざるを得ず，多大な時間を要するという課題がある（図 2）．

図 2　直感的動画検索システムにおける課題

3.3　本研究のアプローチ

直感的動画検索システムにおける課題を解決するため，我々はニコニコ動画に投稿されたコメントを分析することによって印象シーンの出現箇所を特定する手法の開発を進めてきた．図 3 に，本手法による時系列に沿った印象分析のイメージを示す．このように，動画内の印象シーンの出現箇所を分析することで，ユーザは動画全体を視聴することなく，自身が真に見たいと感じる箇所を容易に見つけ出すことが可能となる．本稿では，このニコニコ動画のデータを用いた時系列印象分析手法において，モデル構築手法や分析アルゴリズムにおけるパラメータ設定を見直し，印象シーンの検出精度のさらなる向上を図る．

時系列分析の対象とする印象項目については，直感的動画検索システムの評価項目である「エモい」「きゅん」「映え」に加え，新たに「笑い」を採用した計 4 項目とする．「笑い」を採用した理由は以下の 2 点である．第一に，短文コメントが中心であるニコニコ動画の特性上，文脈理解を要する「じわる」の識別は困難であり，同プラットフォームで支配的な「笑い」が代替として適切であると判断したためである．第二に，シーン検索の需要として，不快感を与える恐れのある「イラつく」などのネガティブな印象よりも，ユーザ体験を向上させるポジティブな印象シーンの特定を優先すべきと判断したためである．

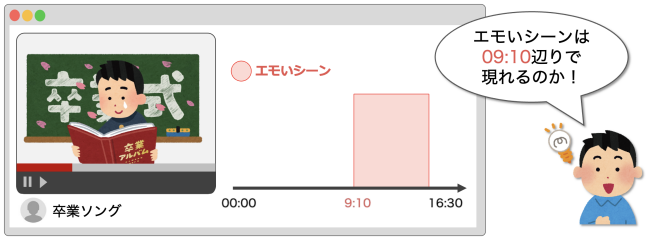

図 3　時系列に沿った印象分析イメージ

4　時系列分析のためのモデル構築手法の改良

本章では，ニコニコ動画のコメント分析に用いるモデルの構築手法の改良について述べる．本研究では，以下の手順で適切なモデルを決定する．まず，学習（ファインチューニング）および評価に使用するデータセットを構築する．次に，そのデータに対する前処理，およびグリッドサーチによるハイパーパラメータの探索を行う．最後に，事前学習済みモデルの種類による比較実験を行い，本タスクに最も適したモデルを選定する．

4.1　データセットの構築

モデルの学習 (ファインチューニング) および評価を行うため，以下の手順でデータセットの構築および分割を行った．

4.1.1　データの収集

ニコニコ動画のコメントから，分析対象の 4 項目（エモい，きゅん，笑い，映え）について，各項目に該当するコメント（正例）と該当しないコメント（負例）をそれぞれ 200 件ずつ，計 400 件収集した．なお，正解ラベルの付与は人手により行った．

4.1.2　データの分割

収集したデータセットは，モデルの学習用と評価用に 7:3 の割合で分割した．具体的には，全 400 件のうち 280 件（正例 140 件，負例 140 件）を学習データ，残りの 120 件（正例 60 件，負例 60 件）をテストデータとして使用した．後述するハイパーパラメータの探索には学習データ（280 件）を使用し，最終的なモデルの精度評価（F 値の算出）にはテストデータ（120 件）を使用する．

4.2　分類精度向上のためのアプローチ

これまでの直感的動画検索システム [1] で使用していた分類モデル（YouTube データを用いて BERT をファインチューニングした既存モデル．以下，YouTube モデル）は，言語的特徴の差異によりニコニコ動画には適さない．そこで我々は，前節で構築したニコニコ動画のデータセットを用いて BERT をファインチューニングし，新たな分類モデル（以下，ニコニコモデル（ベースライン））を構築した．図 4 に，ニコニコ動画のコメント（テストデータ 120 件）に対して YouTube モデルとニコニコモデルを適用した場合の F 値の比較を示す．本節では，図 4 におけるニコニコモデルの精度をさらに向上させるため，前処理の導入およびハイパーパラメータ探索について述べる．

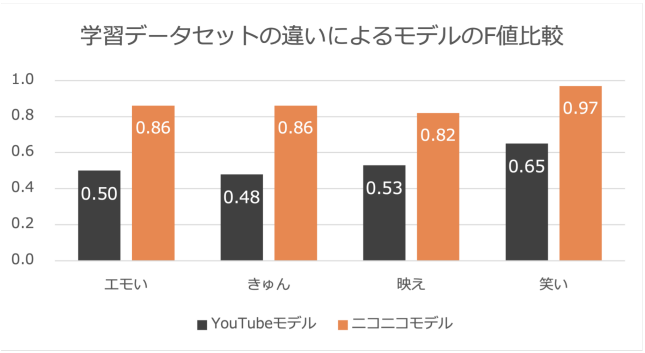

図 4　学習データセットの違いによるモデルの F 値比較

4.2.1　前　処　理

ニコニコ動画に投稿されているコメントには特殊記号や絵文字などのノイズが多く含まれており，これらが学習の妨げとなる可能性がある．そこで，学習データに対し以下のテキスト前処理を導入することで，分類精度の向上を図る．

- **テキストの正規化**: Unicode 正規化（NFKC）を行い，全角・半角の揺らぎや異体字を統一した．また，カタカナをひらがなに変換し，表記揺れを吸収．

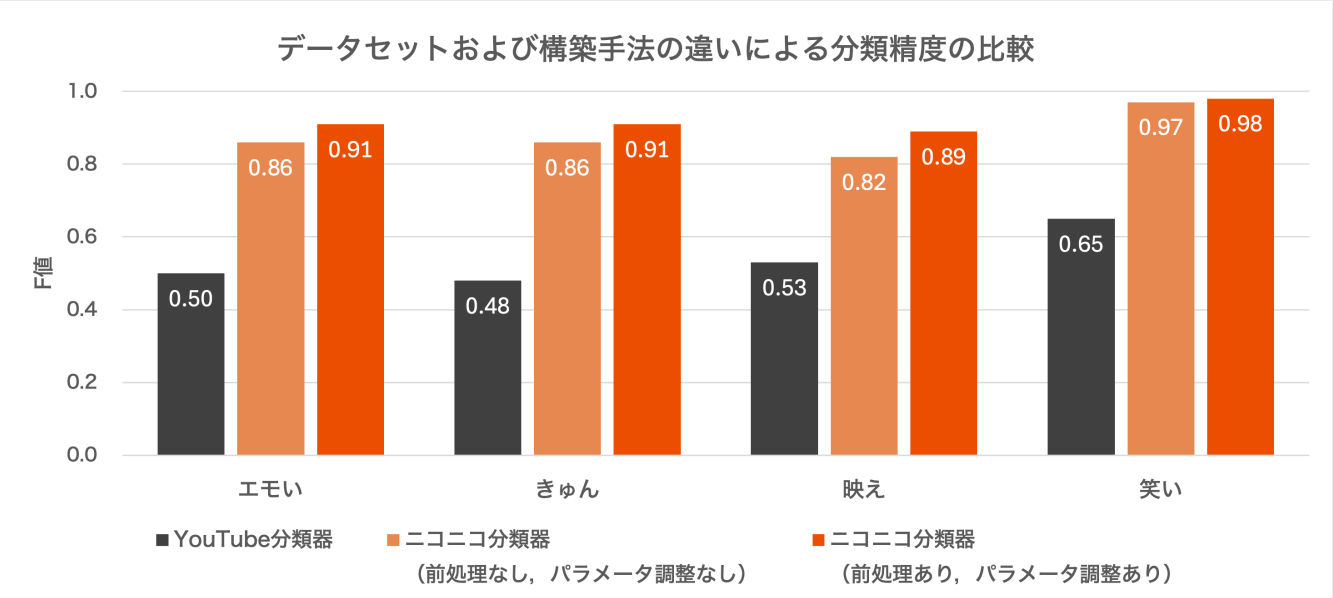

図 5　学習データセットおよび構築手法の違いによる F 値比較

- **非言語情報の除去**: URL，アンカー（>> 1 など），矢印記号（← など），および文脈理解に寄与しない特殊記号を正規表現により除去．
- **表記揺れの統一**: 「草」を「w」，「うまそ」を「美味しそう」に置換するなど，ネットスラングや口語表現を標準的な語句や記号に統一．
- **重複表現の短縮**: 頻繁に使用される文字の連続（例:「wwww」「！！！！」）を，単一の文字（「w」「！」）に置換・短縮することで，語彙のスパース性を低減．

4.2.2　ハイパーパラメータの探索

モデルの性能を最大化するため，ファインチューニング時のハイパーパラメータについて，グリッドサーチによる探索を行う．探索には前節で分割した学習データ（280 件）を用いてファインチューニングを行い，以下の探索範囲で最良となるパラメータを選定する．

- **エポック数**: $\{4\}$
- **バッチサイズ**: $\{16, 32\}$
- **学習率**: $\{1.0 \times 10^{-5}, 2.0 \times 10^{-5}, 3.0 \times 10^{-5}, 5.0 \times 10^{-5}\}$

探索の結果，各印象項目において精度（F 値）が最も高かったパラメータの組み合わせを表 1 に示す．以降の実験においては，これらのパラメータを使用する．

表 1　グリッドサーチにより選定されたハイパーパラメータ

評価項目	学習率	バッチサイズ
エモい	5.0×10^{-5}	16
きゅん	5.0×10^{-5}	16
映え	5.0×10^{-5}	16
笑い	5.0×10^{-5}	16

4.2.3　事前学習済みモデルの選定

近年，自然言語処理分野では，BERT の登場以降，そのアーキテクチャや学習手法を改善し，さらなる性能向上を図った派生モデルが数多く提案されている．そこで本研究では，ニコニコ動画の印象シーン分析において最も高い精度が得られるモデルを選定するため，BERT およびその改良モデルである以下の 3 種類のモデルを対象に比較検討を行う．

1. **BERT** [15]: 文脈を双方向から学習する標準的なモデル．
2. **RoBERTa** [16]: BERT の学習手法を最適化し，より多くのデータで学習させたモデル．
3. **DeBERTa** [17]: 注意機構（Attention Mechanism）を改良し，位置情報の扱いを強化したモデル．

4.3　改良手法の有効性検証（比較実験）

構築するモデルの仕様を決定するため，以下の 2 段階の比較実験を行った．なお，評価にはテストデータ（120 件）を使用する．

4.3.1　実験 1：学習データと構築手法の影響

まず，標準的な BERT モデルを使用し，学習データの違いおよび構築手法（前処理・パラメータ調整の有無）によるテストデータの分類精度（F 値）への影響を検証した．比較対象は以下の 3 通りである．

1. **YouTube モデル**: YouTube データを用いて BERT をファインチューニングしたモデル．
2. **ニコニコモデル（ベースライン）**: 生のニコニコデータで BERT をファインチューニングしたモデル．前処理なし，ハイパーパラメータは標準的な固定値を使用．
3. **改良型ニコニコモデル（提案手法）**: 前処理済みのニコニコデータでファインチューニング．ハイパーパラメータはグリッドサーチにより選定された値を使用．

結果を図 5 に示す．YouTube モデルと比較してニコニコモデルの分類精度は高く，さらに前処理とパラメータ調整を導入した提案手法が，全ての印象項目において最も高い F 値を示した．これにより，ニコニコ動画に特化した前処理およびパラメータ探索の有効性が確認された．

4.3.2 実験 2：事前学習済みモデルの比較

次に，実験 1 で最も精度の高かった「提案手法（前処理あり・パラメータ調整あり）」の設定を用い，3 つの事前学習済みモデル（BERT，RoBERTa，DeBERTa）の精度比較を行った．結果を表 2 に示す．

表 2　事前学習済みモデルの違いによる F 値の比較

モデル名	エモい	きゅん	映え	笑い	平均
BERT	**0.91**	**0.91**	**0.89**	**0.92**	**0.92**
RoBERTa	0.81	0.79	0.82	0.85	0.76
DeBERTa	0.81	0.79	0.82	0.98	0.85

実験の結果，本タスクにおいては BERT が最も高い F 値（平均 0.92）を記録し，比較した 3 モデルの中で最も良好な結果を示した．

この結果に対する考察を述べる．一般に，RoBERTa や DeBERTa は BERT の改良モデルであり，多くのタスクで BERT を上回る性能を持つとされる．しかし本研究においては，ニコニコ動画特有の「極めて短い文」や「特異な表現（スラング等）」が多用されるデータセットを用いたため，複雑な学習機構を持つ改良モデルよりも，標準的な BERT の方がデータの特性に対して過学習を起こしにくく，安定した性能を発揮したと考えられる．

以上の比較実験により，本研究における時系列分析には，「前処理およびパラメータ探索を経たニコニコ動画データ」を用いてファインチューニングを行った「BERT」を採用することとした．次章では，このモデルを用いた具体的な分析手法について述べる．

5　印象の時系列分析手法

本章では，前章で選定した「前処理あり BERT モデル」を用い，動画内の印象シーンを時系列に沿って特定する手法について述べる．

5.1　コメントごとの分類確率算出

本節では，4 章で構築したモデルの使用方法について述べる．4 章で採用された BERT モデルは，対象動画内のコメントそれぞれに対して「各印象項目に該当する確率」を算出する．図 6 に，印象項目「エモい」における確率の算出イメージを示す．

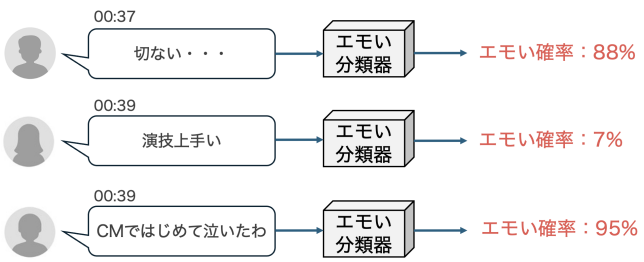

図 6　モデルによるコメントごとの確率算出

5.2　印象シーンの特定アルゴリズム

単一のコメントの確率のみに依存した判定では，文脈と無関係な単発的なコメントやノイズの影響を受けやすく，誤検出の要因となる．そこで本研究では，時系列上の一定区間（ウィンドウ）におけるコメント密度と，その中での印象コメントの含有率（割合）を考慮したシーン判定アルゴリズムを提案する．

本アルゴリズムでは，以下の 4 つの変数を定義し，図 7 に示すフローチャートに従って判定を行う．

1. **印象コメントの判定** (S_{th})：モデルが出力する確率が閾値 S_{th} 以上のコメントを，当該印象（例：エモい）を含む「印象コメント」と定義する．
2. **ウィンドウの設定** (T_{win})：動画全体を秒数 T_{win} ごとのウィンドウに分割し，ここで，分割された i 番目のウィンドウを W_i と定義する．
3. **コメント密度の判定** ($N_{\min}$)：ウィンドウ内に含まれる「印象コメント」の数が $N_{\min}$ 未満の場合，特定に必要な情報量が不足していると判断し，当該ウィンドウは「印象シーンではない」と判定（棄却）する．
4. **含有率によるシーン判定** (R_{th})：ウィンドウ内の全コメント数に対し，「印象コメント」が占める割合を算出する．この含有率が閾値 R_{th} 以上である場合，当該ウィンドウを最終的に「印象シーン」として特定する．

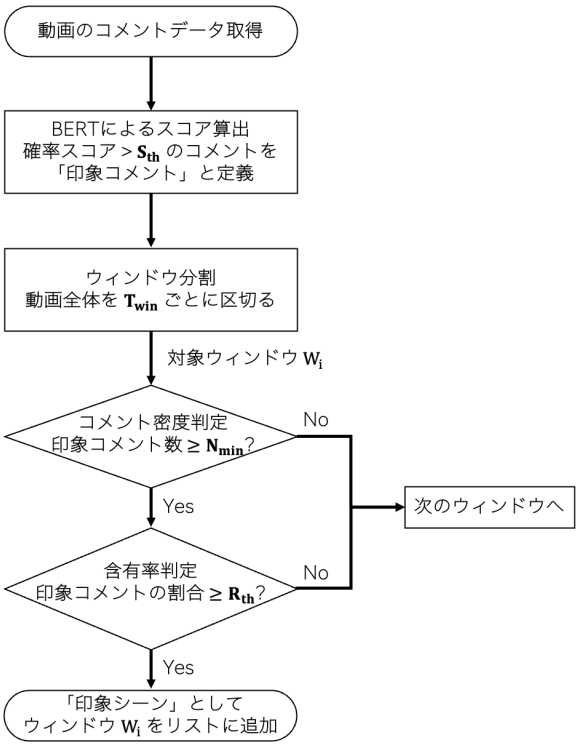

図 7　印象シーン検出の処理フロー

6 実データによる検証実験

6.1 実験の目的と評価指標

本章では，提案したシーン特定アルゴリズムの有効性を検証し，適切なパラメータ設定を決定するための実験について述べる．本実験における評価指標として，一般的な F_1 スコアではなく，$F_{0.5}$ スコアを採用する．

$F_{0.5}$ スコアは適合率（Precision）と再現率（Recall）を用いて以下の式 (1) で定義される．

$$F_{0.5} = (1 + 0.5^2) \cdot \frac{\text{Precision} \cdot \text{Recall}}{(0.5^2 \cdot \text{Precision}) + \text{Recall}} \qquad (1)$$

動画検索や推薦システムにおいて，実際には印象シーンではない箇所を誤って検出することは，ユーザに「期待外れ」という負の感情を与え，システムの信頼性を損なう要因となる．そのため本研究では，網羅性（再現率：Recall）よりも，検出したシーンが本当に正しいか（適合率：Precision）を重視する方針をとり，適合率に重みを置いた $F_{0.5}$ 値を指標として用いる．

6.2 データセットと正解定義

実験データとして,「エモい」「きゅん」「笑い」「映え」の各印象項目につき，ニコニコ動画から人手により当該印象シーンを含むと判断した動画を各 10 本（計 40 本）選定し，使用する．これらの動画に対し，評価者が実際に視聴を行い，印象を感じる区間を手動で選定したものを正解区間とした．評価は動画の再生時間 1 秒単位で行い，提案手法により検出された区間が正解データと重複しているか否かを判定する．

本研究では，これらの判定に基づき，適合率（Precision）と再現率（Recall）を以下の式 (2) および式 (3) で定義する．

$$\text{Precision} = \frac{\text{正しく検出されたシーンの総時間 [秒]}}{\text{システムが検出したシーンの総時間 [秒]}} \qquad (2)$$

$$\text{Recall} = \frac{\text{正しく検出されたシーンの総時間 [秒]}}{\text{人手で作成した正解シーンの総時間 [秒]}} \qquad (3)$$

6.3 実験方法：パラメータの探索と評価

6.3.1 パラメータの探索と評価に使用するデータセット

6.2 で述べたデータセットを各印象項目ごとにパラメータ探索用 5 本，評価用 5 本に分割する．

6.3.2 パラメータの探索

前節で述べた，各印象項目ごとのパラメータ探索用の動画 5 本，計 20 本を使用して適切なパラメータの組み合わせ探索を行う．パラメータ探索範囲は表 3 に示す通りである．

各印象項目において，パラメータ探索用動画 5 本の平均 $F_{0.5}$ 値が最大となるパラメータの組み合わせを特定した．特定されたパラメータと，その際の平均 $F_{0.5}$ 値を表 4 に示す．

表 3　パラメータ探索における候補値一覧

パラメータ	探索候補値
確率閾値 (S_{th})	$0.5, 0.7, 0.9$
ウィンドウ幅 (T_{win})	$3, 5$ (秒)
印象コメントの最低数 ($N_{\min}$)	$1, 3, 5, 7, 9$ (件)
含有率閾値 (R_{th})	$0.1, 0.3, 0.5, 0.7$

表 4　グリッドサーチにより選定されたパラメータと $F_{0.5}$ 値

評価項目	T_{win}	S_{th}	$N_{\min}$	R_{th}	$F_{0.5}$ 値
エモい	5	0.5	5	0.3	0.34
きゅん	3	0.7	3	0.3	0.64
映え	5	0.7	5	0.3	0.26
笑い	5	0.7	5	0.1	0.35

6.4 実験結果と考察

パラメータ探索による効果を検証するため，評価用動画（各印象項目ごとに 5 本）に対して，探索前（ベースライン）と探索後の設定で印象シーンの検出精度の比較を行った．探索前のパラメータは（$S_{th} = 0.5, T_{win} = 5, N_{\min} = 3, R_{th} = 0.6$）となっている．パラメータ調整前後の比較結果を表 5 に示す．

表 5　初期パラメータと選定パラメータ適用後の精度比較

評価項目	探索前 $F_{0.5}$	探索後 $F_{0.5}$	改善幅
エモい	0.02	0.29	+0.27
きゅん	0.05	0.33	+0.28
映え	0.02	0.14	+0.12
笑い	0.20	0.51	+0.31

実験の結果，選定されたパラメータを適用することで，すべての印象項目において $F_{0.5}$ 値の大幅な向上が確認された．しかし，「映え」に関しては改善は見られたものの，他の項目と比較して低いスコア（0.14）に留まった．この要因として，学習データと実際のシーンにおける言語的特徴の乖離が挙げられる．モデルのファインチューニング時には「綺麗」や「美しい」といった記述的なコメントを正例として用いたが，実際の「映えシーン」では「うおおお」「すげええ」といった感嘆詞的なコメントが支配的であった．この語彙特性の不一致により，モデルがシーンの特徴を十分に捉えきれなかったと推察される．

以上の結果より，本手法におけるパラメータ探索の有効性が示されたが，印象項目ごとのコメント特性に応じた更なる学習データの拡充やモデル改良が必要であることが示唆された．

7 おわりに

本稿では，直感的動画検索システムの機能拡張として，ニコニコ動画のコメントデータを用いた印象シーンの時系列分析手法の改良について述べた．分析の信頼性を高めるため，BERT を用いたモデルの導入と前処理の適用を行い，従来手法よりも高精度な分類を実現した．さらに，シーン検出においては適合率を重視した $F_{0.5}$ 値を評価指標とし，パラメータの探索を行った．実データを用いた評価実験の結果，パラメータの調整により，ノイズによる誤検出を抑制し，シーン検出精度を向上でき

ることを確認した．特に「笑い」などの項目では高い改善効果が得られた一方，「映え」のように言語的特徴の乖離が課題となる項目も明らかとなった．

今後は，感嘆表現などの多様なコメントパターンに対応できるよう学習データを拡充するとともに，直感的なユーザインタフェースの実装について検討を進める．

謝　　辞

本研究の一部は，JSPS 科研費（課題番号：22K12281）および京都産業大学先端科学技術研究所（人間情報学研究センター）共同研究プロジェクト（M2301）の助成を受けたものである．また，国立情報学研究所の IDR データセット提供サービスにより株式会社ドワンゴから提供を受けた「ニコニコ動画コメント等データ」を利用した．ここに記して謝意を表す．

文　　献

[1] 鴇田和士, 上田真由美, 中島伸介. 動画に対する評価項目別スコアを用いた直感的動画検索システム．第 14 回データ工学と情報マネジメントに関するフォーラム (DEIM2022), No. C21-5, 2022.

[2] Akito Sasahara, Asaka Cheng Lan, Da Li, Fumiya Yamaguchi, Mayumi Ueda, and Shinsuke Nakajima. Leveraging comment data for enhanced content discovery through time series analysis of impressive scenes in videos. In *Proceedings of the 12th International Conference on Behavioural and Social Computing (BESC 2025)*, Hong Kong, China, October 2025. Springer.

[3] Katja Hofmann, Shimon Whiteson, and Maarten de Rijke. A probabilistic method for inferring preferences from clicks. In *Proceedings of the 20th ACM Conference on Information and Knowledge Management (CIKM 2011)*, pp. 249–258, Glasgow, United Kingdom, October 2011.

[4] 加藤誠, 山本岳洋, 真鍋知博, 西田成臣, 藤田澄男. コミュニティ q&a サイトにおける質問検索システムの大規模オンライン評価. 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), pp. G2–2, 2018.

[5] S. Gheewala, A. Sharma, R. Patel, M. Tanaka, L. Zhang, and D. Kim. Exploiting deep transformer models in textual review based recommender systems. *Expert Systems with Applications*, Vol. 235, , 2024.

[6] K. Sharma, Y. Liu, M. Gupta, T. Nguyen, and P. Roy. A survey of graph neural networks for social recommender systems. *ACM Computing Surveys*, Vol. 56, No. 10, pp. 1–34, 2024.

[7] 村上直至, 伊東栄典. 動画コンテンツの視聴者コメントに基づくランキングとその評価. 第 4 回データ工学と情報マネジメントに関するフォーラム (DEIM2012), No. F8-3, 2012.

[8] 中村聡史ほか. ソーシャルアノテーションに基づく動画検索手法. 第 1 回データ工学と情報マネジメントに関するフォーラム (DEIM2009), pp. D6–1, 2009.

[9] Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. How useful are your comments? analyzing and predicting youtube comments and comment ratings. In *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*, pp. 891–900, 2010.

[10] Mohan Sai Dinesh Boddapati, Madhavi Sai Chatradi, Sridevi Bonthu, and Abhinav Dayal. Youtube comment analysis using lexicon based techniques. In *Cognitive Computing and Cyber Physical Systems*, pp. 76–85. Springer Nature Switzerland, 2023.

[11] Michal Muszynski, Leimin Tian, Catherine Lai, Johanna D. Moore, Theodoros Kostoulas, Patrizia Lombardo, Thierry Pun, and Guillaume Chanel. Recognizing induced emotions of movie audiences from multimodal information. *IEEE Transactions on Affective Computing*, Vol. 12, No. 1, pp. 36–52, 2021.

[12] 尾崎遼太郎, 佐々木史織, 清木康. 動画とライブ配信のコメントデータを対象としたリアルタイム時系列感性変化計量・可視化・検索システム．第 11 回データ工学と情報マネジメントに関するフォーラム (DEIM2019), pp. C4–1, 2019.

[13] Asutosh Rath, B. Hridaya, D. Vimala, and J. George. Multilingual sentiment analysis of youtube live stream using machine translation and transformer in nlp. In *2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT)*, pp. 1–5, Pune, India, 2022.

[14] M. S. Barakat, C. H. Ritz, and D. A. Stirling. Temporal sentiment detection for user generated video product reviews. In *2013 13th International Symposium on Communications and Information Technologies (ISCIT)*, pp. 580–584, 2013.

[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[16] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

[17] Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. Deberta: Decoding-enhanced bert with disentangled attention, 2021.